Lecture 19: Cramér-Rao lower bound

17 March 2016

1 Cramér-Rao Lower Bound (CRLB)(Continued from Lecture 18)

Let $\mathbf{X} \sim f(x|\theta), \theta \in \mathbb{R}$. Suppose:

$$\frac{d}{d\theta}\mathbb{E}[h(\mathbf{X})] = \int \frac{\partial}{\partial\theta}[h(\mathbf{x})f(\mathbf{x}|\theta]d\mathbf{x}. \tag{1}$$

(domain of the above integration does not depend on θ) for $h(\mathbf{x}) = W(\mathbf{x})$, where $W(\mathbf{x})$ is an estimator. Now for $h = 1$,

$$Var_\theta(W) \geq \frac{(\frac{d}{d\theta}\mathbb{E}_\theta[W])^2}{\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log(f(x|\theta)))^2]}. \tag{2}$$

The Sufficient condition required for (1),

$$\frac{d}{d\theta}\int_\Gamma f(x,\theta)d\mathbf{x} = \int_\Gamma \frac{\partial}{\partial\theta}f(x,\theta)d\mathbf{x}, \tag{3}$$

holds when either,

1. $\Gamma \subseteq \mathbb{R}^d$ is compact (closed and bounded), and $\frac{\partial}{\partial\theta}f(x,\theta)$ is continuous over all x, or

2. $\int_\Gamma |\frac{\partial}{\partial\theta}f(x,\theta)|d\mathbf{x} < \infty$.

1.1 Multivariate CRLB

Let $\mathbf{X} \sim f(x|\theta), \theta \in \Theta, \Theta \subseteq \mathbb{R}^d$ and $W(\mathbf{X})$ is an estimator,

$$\frac{\partial}{\partial\theta_i}\mathbb{E}_{\theta_i}[h(\mathbf{X})] = \int \frac{\partial}{\partial\theta_i}[h(\mathbf{x})f(\mathbf{x}|\theta)]d\mathbf{x}, \quad \forall i \in [d] \tag{4}$$

for $h = W$. Then,

$$Var_\theta(W) \geq (\nabla_\theta \psi)^T I(\theta)^{-1} (\nabla_\theta \psi) \tag{5}$$

where, $\psi(\theta) = \mathbb{E}_\theta[W(X)]$, and $I(\theta)_{i,j} = \mathbb{E}_\theta[\frac{\partial}{\partial \theta_i} \log(f(\mathbf{x}|\theta)) \frac{\partial}{\partial \theta_j} \log(f(\mathbf{x}|\theta))]$.

Corollary 1.1 (CRLB for iid Samples). *If the assumptions of Cramer-Rao inequality are satisfied and, additionally, if $X_1, \ldots, X_n$ be iid $\sim f(x|\theta)$, then*

$$Var_\theta W(\mathbf{X}) \geq \frac{(\frac{d}{d\theta}\mathbb{E}_\theta W(\mathbf{X}))^2}{n\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log(f(X|\theta)))^2]}.$$

Proof. To prove this we only need to show

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(X_i|\theta)\right)^2\right] = n\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right]. \tag{6}$$

Since $X_1, \ldots, X_n$ are independent,

$$\begin{aligned}\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(X_i|\theta)\right)^2\right] &= \mathbb{E}_\theta\left[\left(\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f(X_i|\theta)\right)^2\right], \\ &= \sum_{i=1}^n \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\right)^2\right] + \\ &\quad \sum_{i\neq j}\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X_i|\theta))\frac{\partial}{\partial\theta}\log f(X_j|\theta)\right].\end{aligned}$$

For $i \neq j$, we have

$$\begin{aligned}\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X_i|\theta)\frac{\partial}{\partial\theta}\log f(X_j|\theta)\right] &= \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X_i|\theta)\right]\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X_j|\theta)\right], \\ &= 0.\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(X_i|\theta)\right)^2\right] &= \sum_{i=1}^n \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\right)^2\right] \\ &= n\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right]. \end{aligned} \tag{7}$$

□

The quantity $\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log(f(x|\theta))^2]$ is called the Fisher information of the sample and $\frac{\partial}{\partial\theta}\log f(X|\theta)$ is called the *score function.*

Lemma 1.2. *If $f(x|\theta), \theta \in \Theta$ satisfies*

$$\frac{d}{d\theta}\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X|\theta)\right] = \int \frac{\partial^2}{\partial\theta^2}[\log f(x|\theta).f(x|\theta)]dx,$$

then, the Fisher information

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log(f(X|\theta)\right)^2\right] = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log(f(X|\theta))\right]. \tag{8}$$

Note 1. This is true for Exponential family distribution

$$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right).$$

Example 1.3 (Poisson CRLB). Let $X_1, \ldots, X_n$ be iid Poisson(λ), $\lambda > 0$ and $W(X)$ is an estimator of λ.

$$\frac{d}{d\lambda}\mathbb{E}_\lambda[W] = 1.$$

$$\begin{aligned}
\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(X_i|\theta)\right)^2\right] &= -n\mathbb{E}_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log\left(\frac{e^{-\lambda}\lambda^X}{X!}\right)\right], \\
&= -n\mathbb{E}_\lambda\left[\frac{\partial^2}{\partial\lambda^2}(-\lambda + X\log\lambda - \log X!)\right], \\
&= -n\mathbb{E}_\lambda\left[\frac{-X}{\lambda^2}\right], \\
&= \frac{n}{\lambda}.
\end{aligned}$$

Hence for any unbiased estimator W, of λ, we must have

$$Var_\lambda[W] \geq \frac{\lambda}{n}. \tag{9}$$

Since,

$$Var_\lambda[\bar{X}] = Var_\lambda[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{\lambda}{n}, \tag{10}$$

$\bar{X}$ is a best unbiased estimator of λ.

Example 1.4 (Normal CRLB). Let $X_1, \ldots, X_n$ be Normal iid $\mathcal{N}(\mu, \sigma^2)$

1. Unbiased estimation of μ: $\bar{X}$ meets CRLB
2. Unbiased estimation of Variance(σ^2): Normal pdf satisfies the assumption of the Cramer-Rao theorem and lemma, so we have

$$\begin{aligned}\frac{\partial^2}{\partial(\sigma^2)^2}\log\{\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}\} &= \frac{\partial^2}{\partial(\sigma^2)^2}\{\log(\frac{1}{\sqrt{\sigma^2}}) - \frac{(x-\mu)^2}{2\sigma^2}\},\\ &= \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6},\end{aligned}$$

and

$$\begin{aligned}-\mathbb{E}_(\mu, \sigma^2)\left[\frac{\partial^2}{\partial(\sigma^2)^2}\log(f(X|(\mu,\sigma^2))\right] &= -\mathbb{E}_(\mu, \sigma^2)\left[\frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}\right],\\ &= \frac{-1}{2\sigma^4} + \frac{1}{\sigma^4},\\ &= \frac{1}{2\sigma^4}.\end{aligned}$$

Thus CRLB for any unbiased estimator W of σ^2 is

$$Var[W] \geq \frac{2\sigma^4}{n}. \tag{11}$$

We saw earlier that

$$Var[S^2] = \frac{2\sigma^4}{n-1}.$$

Hence, S^2 does not attain the Cramer-Rao Lower Bound.

Corollary 1.5 (Attainment of CRLB). *par Let* $\mathbf{X} \sim f(x|\theta)$. $W(X)$ *be an unbiased estimator of* $g(\theta)$ *and* f, W *satisfy the condition of CRLB Hypothesis. Then* W *attains the CRLB if and only if* $\exists$ *a function* $a(\theta)$ *such that*

$$\frac{\partial}{\partial\theta}\log(f(\mathbf{x}|\theta)) = a(\theta)[W(\mathbf{x}) - g(\theta)], \quad \forall\theta. \tag{12}$$

(i.e. the Score function is a affine function of the estimator)

Proof. For any two random variables A and B

$$Cov(A,B)^2 \leq Var(A)Var(B) \tag{13}$$

$$\mathbb{E}[AB]^2 \leq \mathbb{E}[A^2]\mathbb{E}[B^2] \tag{14}$$

We can have equality in CRLB if and only if

$$(A - \mathbb{E}[A]) = \alpha(B - \mathbb{E}[B]), \tag{15}$$

where $A = \frac{\partial}{\partial\theta}\log(f(\mathbf{x}|\theta))$, and $B = W$.

Recalling that $\mathbb{E}_\theta[W] = g(\theta)$ and $\mathbb{E}_\theta[\frac{\partial}{\partial\theta}\log f(\mathbf{x}|\theta)] = 0$,

$$\frac{\partial}{\partial\theta}\log(f(\mathbf{x}|\theta)) = \alpha(\theta)[W - g(\theta)], \quad \forall\theta. \tag{16}$$

□

Example 1.6 (Application). Let $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, then we have

$$L(\mu, \sigma^2|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\} \tag{17}$$

and hence

$$\begin{aligned} \frac{\partial}{\partial\sigma^2}\log L(\mu, \sigma^2|\mathbf{x}) &= -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^4}, \\ &= \frac{n}{2\sigma^4}\{\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{n} - \sigma^2\}. \end{aligned} \tag{18}$$

Thus taking $a(\sigma^2) = \frac{n}{2\sigma^4}$ shows that the best estimator of σ^2 is $\sum_{i=1}^{n}(x_i-\mu)^2/n$, which is calculable only if μ is known. If μ is unknown, the bound can not be attained.

2 Unbiased Estimation & Sufficient Statistics

Recall that a sufficient statistics $T(X)$, for θ, where $X \sim f(x|\theta)$ is one for which $P(X|T(X))$ does not depend on θ. Thus sufficient statistics help in finding low variance estimator.

Theorem 2.1 (Rao-Blackwell). *Let $X \sim f(x|\theta)$,$\theta \in \Theta$ and $W(X)$ be any unbiased estimator of $g(\theta)$ and T be a sufficient statistics of θ. Define the estimator $\phi = \mathbb{E}[W|T]$ (i.e. $\phi(x) = \mathbb{E}[W(X)|T(X) = T(x)]$). Then,*

1. $\mathbb{E}_\theta[\phi] = g(\theta), \quad \forall\theta$
2. $Var_\theta[\phi] \leq Var_\theta[W], \quad \forall\theta.$

That is ϕ is a uniformly better unbiased estimator of $g(\theta)$ (in the sense of MSE).

Proof.

$$\begin{aligned} g(\theta) &= \mathbb{E}_\theta[W], \\ &= \mathbb{E}_\theta[\mathbb{E}_\theta[W|T]]. \\ &= \mathbb{E}_\theta[\phi]. \end{aligned} \tag{19}$$

So ϕ is unbiased for $g(\theta)$.

Lemma 2.2. *For two random variables $X\&Y$*

$$Var[X] = Vax[\mathbb{E}[X|Y]] + \mathbb{E}[Var[X|Y]]. \tag{20}$$

Using the above lemma

$$\begin{aligned} Var_\theta[W] &= Var_\theta[\mathbb{E}[W|T]] + \mathbb{E}_\theta[Var_\theta[W|T]], \\ &= Var_\theta[\phi] + \mathbb{E}_\theta[Var_\theta[W|T]], \\ &\geq Var_\theta[\phi], \quad (Var_\theta[W|T] \geq 0). \end{aligned} \tag{21}$$

Hence ϕ is uniformly better then W. □